TR-214

WHOI
DOCUMENT
COLLECTION

TECHNICAL REPORT

FITTING A SET OF STRAIGHT LINES TO A DIGITAL BT PROFILE

APRIL 1969

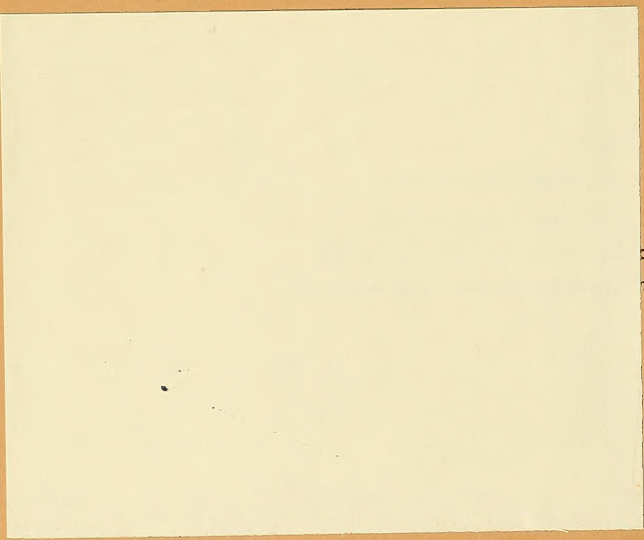
GC
1
.T43
no.TR-214

NAVAL OCEANOGRAPHIC OFFICE
WASHINGTON, D. C. 20390
Price 60 cents

SHOUT
SHARE
IT

ABSTRACT

A "three level tolerance algorithm" for fitting sets of straight lines to points digitized at regular depth intervals from a BT profile is discussed. The application of this algorithm to the reduction of other profile data similarly digitized is discussed.



FOREWORD

The rapid increase in the rate of growth of oceanographic data files has stimulated an awareness of the problems related to the management of large data masses and has motivated attempts to devise new techniques for resolving these problems effectively. Accordingly, the National Data Program for the Marine Environment is concerned with the overall problem as it exists now and as it is projected into the future. To deal with the immediate problem of processing existing data with reasonable economy in a reasonable time frame, the Marine Sciences and the Research and Development Departments of the Naval Oceanographic Office and the National Oceanographic Data Center are jointly developing an improved system for the management of ocean data files. This system, which has been referred to as a "live atlas", is conceived as a tool for providing the oceanographer a quick response computer feed-back data analysis capability, whereby questions formulated in response to a product display may be resolved immediately with the appropriate selection of subsequent displays.

It is essential that the live atlas provide the most extensive, reliable and flexible data base that can be achieved within the limited reserves of quick access computer memory. Accordingly, considerable attention has been given to the problems of improving data quality and of formatting data for minimum storage requirements and maximum speed processing capabilities. This report documents an algorithm developed as a result of studies for improving upon the quality and manageability of the National Oceanographic Data Center Bathythermograph (BT) file. This algorithm is important, since it may be used to improve the quality and manageability of various profile forms, including expendable BT (XBT) and analog salinity-temperature-depth (STD) data.

T. K. TREADWELL
Captain, U. S. Navy
Commander
U. S. Naval Oceanographic Office

MBL/WHOI
0 5816900 1063 0

DATA LIBRARY
Woods Hole Oceanographic Institution

TABLE OF CONTENTS

Page

INTRODUCTION 1

PROBLEM DEFINITION 1

Preserve Features of Profile Gradient 1

Eliminate Spurious Gradient Irregularities 4

Identify Straight Line Subsets in a Data Point Aggregate 4

PROBLEM SOLUTION 6

Graphical Solution 6

Tolerance a Function of Trace Slope 6

Trial and Error Regression Method 8

Intersection 10

Smoothing 12

Width of the "Thin" Stylus Line 12

Measure of "Appropriateness of Fit" 14

COMMENTS AND CONCLUSIONS 17

Room for Analytical Studies 17

The "Three Level Tolerance Algorithm" 17

Compromise Regression Points 18

Fitting Curvelinear Functions 18

Other Kinds of Profile Data Sets 19

TABLE OF CONTENTS (CONTINUED)

Page

PROFILE REGRESSION PROGRAM 21

Input/Output Data Linkage 21

Subroutines 22

Reduction of Data Memory Volume Requirements 22

Production Time 23

APPENDIX A A-1

FIGURES

1. Sample BT Profiles 2

2. Lines Fit to a Profile Data Set 3

3. Extreme Gradient Error Range 5

4. Comparing Lines of Regression on Straight Line Sets 5

5. Graphical Solution Tolerance 7

6. Intersection 11

7. The "Three Level Tolerance" Algorithm 13

8. Stylus Thickness Value, Regular Profile 15

9. Stylus Thickness Value, Irregular Profile 16

10. Fitting Curvelinear Functions 20

INTRODUCTION

The Bathythermograph (BT) file is one of the more voluminous data files compiled by the National Oceanographic Data Center (NODC). It consists of temperature profile data visually digitized at five meter intervals from ozalid copies of bathythermograph slides (Figure 1). Although the size of an ozalid BT slide copy may vary, a representative scale is about two inches to 300 meters. Accordingly, a profile is digitized at 0.03 inch intervals on the BT graph scale. This fine resolution is ostensibly recommended for representing subtle temperature gradient irregularities. However, certain undesirable effects accompany such fine scale digitization intervals, namely: (1) the temperature gradient, which is calculated using differences between adjacent data points, is overly sensitive to slight variations caused by reader error and truncation, and is accordingly unreliable; (2) a considerable bulk of computer memory is required to store profile data points (as many as sixty data points for the standard scale BT profile).

The computer algorithm described in this paper has been developed to reduce an NODC digital BT profile to a select set of regression lines (Figure 2). The objective of the algorithm is to represent a BT profile with economy and precision (negligible departure from the original BT data points). It has the feature that it provides reasonable estimates of temperature gradient between regression points (points of intersection between neighboring regression lines). This algorithm may, with appropriate modifications, be applied to the reduction of any kind of digital profile data set. For example, it may be used to reduce massive STD profile data sets to manageable proportions without compromising their fine scale descriptions of profile gradient.

PROBLEM DEFINITION

A. Preserve Features of Profile Gradient

A BT profile trace may misrepresent not only real temperature values with depth, but real depth and temperature gradient values as well. The accuracy of a mechanical BT profile curve may be further compromised by mechanical and thermal inertia which causes hysteresis in the BT trace between sensor descent and ascent. Perhaps paradoxically, the hysteresis provides convincing evidence that the BT trace truthfully describes real irregularities in the temperature profile. Even in cases where hysteresis is pronounced, as a general rule, the humps and bumps registered in a trace during sensor descent match humps and bumps registered during sensor ascent (Figure 1(b)). It is a challenge to represent a BT trace digitally without losing the evidence of these subtle humps and bumps.

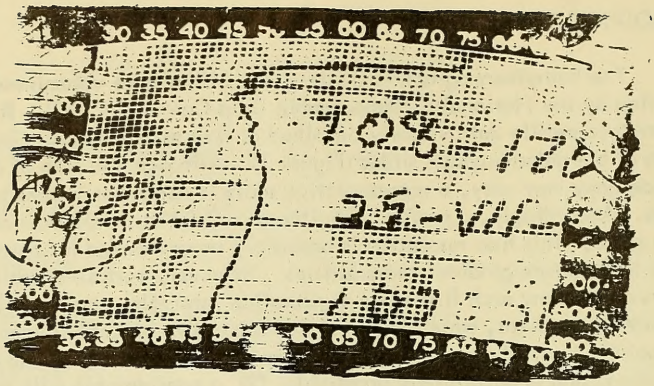


FIGURE 1(a). A TYPICAL OZALID COPY BT PROFILE. NOTE THE SIZE AND SCALE OF THE GRAPH.

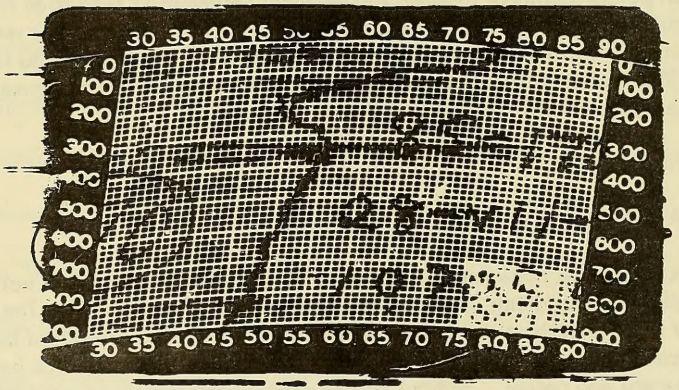


FIGURE 1(b). HYSTERESIS IN A BT PROFILE, CAUSED BY SENSOR MECHANICAL INERTIA. NOTE CORRESPONDENCE BETWEEN "HUMPS AND BUMPS" RECORDED DURING SENSOR DESCENT AND ASCENT.

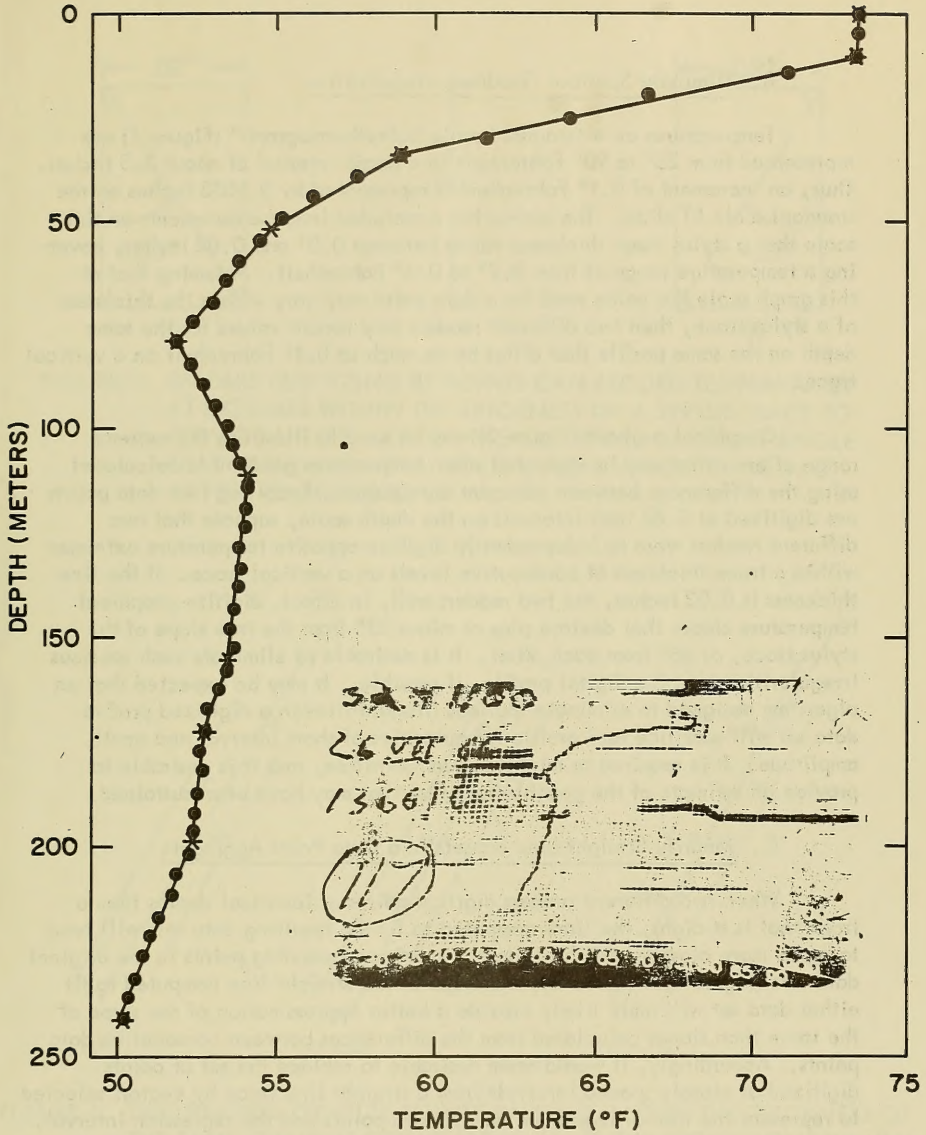


FIGURE 2. THE "SOLUTION PLANE" RECTILINEAR GRID IS DEFINED TO APPROXIMATE THE TEMPERATURE/DEPTH SCALE OF THE STANDARD CURVELINEAR GRID BATHYTHERMOGRAPH. THE BT PROFILE (INSET) WAS VISUALLY DIGITIZED AT FIVE METER OR FIFTEEN FOOT INTERVALS. DATA POINTS (DOTS) ARE PLOTTED ON THE SOLUTION PLANE. "REGRESSION POINTS" (ASTERISKS) WERE CALCULATED USING THE THREE LEVEL TOLERANCE ALGORITHM DESCRIBED IN THIS REPORT.

B. Eliminate Spurious Gradient Irregularities

Temperatures on a "standard scale bathythermograph" (Figure 1) are represented from 28° to 90° Fahrenheit in a graph interval of about 3.3 inches. Thus, an increment of 0.1° Fahrenheit is represented by 0.0053 inches on the standard scale BT slide. The author has concluded from measurements on this scale that a stylus trace thickness varies between 0.01 and 0.02 inches, covering a temperature range of from 0.2° to 0.4° Fahrenheit. Assuming that at this graph scale the value read for a data point may vary within the thickness of a stylus trace, then two different readers may record values for the same depth on the same profile that differ by as much as 0.4° Fahrenheit on a vertical trace.

Graphical methods (Figure 3) may be used to illustrate the extreme range of error that may be expected when temperature gradient is calculated using the differences between adjacent data points. Recalling that data points are digitized at 0.03 inch intervals on the depth scale, suppose that two different readers were to independently digitize opposite temperature extremes within a trace thickness at consecutive levels on a vertical trace. If the line thickness is 0.02 inches, the two readers will, in effect, digitize graphical temperature slopes that deviate plus or minus 33° from the true slope of the stylus trace, or 66° from each other. It is desirable to eliminate such spurious irregularities from the digital profile, if possible. It may be expected that an algorithm designed to eliminate spurious irregularities in a digitized profile data set will sacrifice real profile irregularities of short interval and small amplitude. It is required to minimize such sacrifice, and it is desirable to provide an estimate of the possible sacrifice that may have been sustained.

C. Identify Straight Line Subsets in a Data Point Aggregate

When two different readers digitize points at identical depths from a trace that is straight, the lines computed to fit the resulting data sets will tend to agree more closely, depth for depth, than corresponding points in the original data sets (Figure 4). Moreover, the slope of the straight line computed to fit either data set will more likely provide a better approximation of the slope of the trace than slopes caluclated from the differences between consecutive data points. Accordingly, it would seem desirable to replace the set of points digitized at closely spaced intervals from a straight line trace by vectors selected to represent the line of regression on the data points and the regression interval.

Of course, it happens that most BT profiles cannot be represented by a straight line. However, a profile can be represented by sets of straight lines, each line selected to fit a series of data points that fall within a negligible

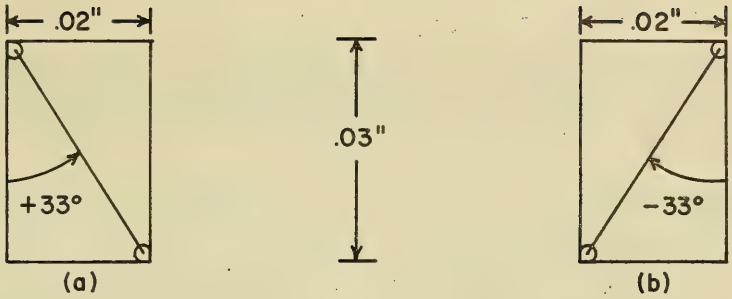


FIGURE 3. READERS DIGITIZING BT POINTS CAN RECORD TEMPERATURES AT EXTREMES WITHIN THE THICKNESS OF A STYLUS TRACE TO PROVIDE AN ESTIMATE OF THE GRAPHICAL GRADIENT ANGLE THAT VARIES FROM THE TRUE GRADIENT ANGLE BY AS MUCH AS PLUS OR MINUS 33°.

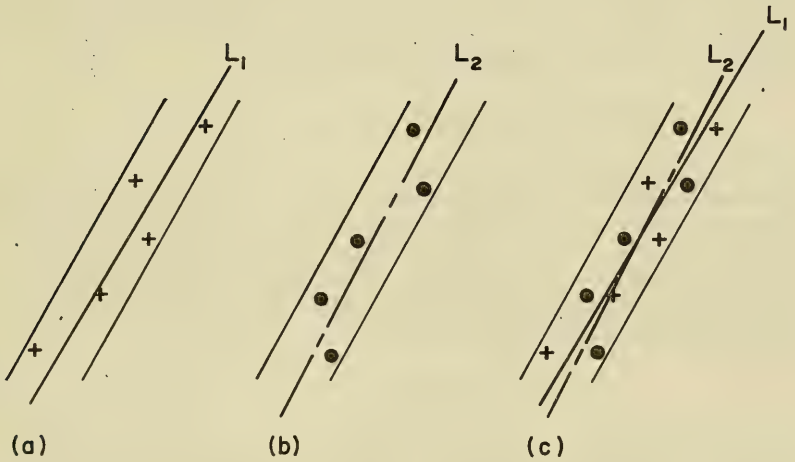


FIGURE 4. TWO READERS DIGITIZE POINTS FALLING WITHIN THE THICKNESS OF THE SAME STRAIGHT STYLUS TRACE. L_1 IS THE REGRESSION LINE ON THE FIRST DATA SET (a), and L_2 IS THE REGRESSION LINE ON THE SECOND DATA SET (b). THE RESULTS ARE SUPERIMPOSED (c) TO ILLUSTRATE THAT THE REGRESSION LINES L_1 AND L_2 WILL TEND TO AGREE MORE CLOSELY, DEPTH FOR DEPTH, THAN THE ORIGINAL DATA POINTS.

distance of a straight line. It remains to determine what constitutes a "negligible distance" of a data point from a straight line, and to exploit methods for dividing a profile data set into subsets each of which is appropriately identified with a straight line.

PROBLEM SOLUTION

A. Graphical Solution

The problem of BT profile reduction, stated in the last section, may be regarded as a graphical problem of fitting lines to points in a plane. The metric of the plane is defined by the dimensions of the standard BT slide scale. It is assumed that a stylus trace on this plane will be of constant thickness independent of its rotation in the plane. And it is assumed appropriate to fit a straight line to points that align within the thickness of a stylus trace.

It may be seen that the depth/temperature grid of the standard BT is not linear (Figure 2). A "solution plane" may be defined in which temperature and depth are related by a rectilinear grid in proportions that approximate the BT grid. The grid of the solution plane is adequate for representing stylus trace thickness as a constant independent of rotation.

B. Tolerance a Function of Trace Slope

Accordingly, a zonal section of the trace will vary in magnitude with the slope of the trace, and points falling within a trace thickness on the plane may be separated by a distance exceeding trace thickness on a zonal section (Figure 5(a)). If a line is computed for least squares fit with points on the trace, then the extreme zonal departure of a point from the straight line may be permitted to vary with the slope of the line. Referring to Figure 5(a), the permitted range of departure is $\pm e$, where

$$e = d/(2 \sin \theta). \tag{1}$$

This is translated to units of depth and temperature as follows. Suppose that one inch on the plane of the BT slide equals k1 intervals on the depth scale and k2 intervals on the temperature scale. Then, referring to Figure 5(b),

$$\sin \theta = 1/\sqrt{1 + \left(\frac{k1}{k2} \cdot \frac{T2 - T1}{k2 - z1}\right)^2} \tag{2}$$

If the permitted range of departure on a vertical trace is expressed in units of temperature, δT, so that $e = \delta T/(2 \sin \theta)$, then the equation (1) may be expressed

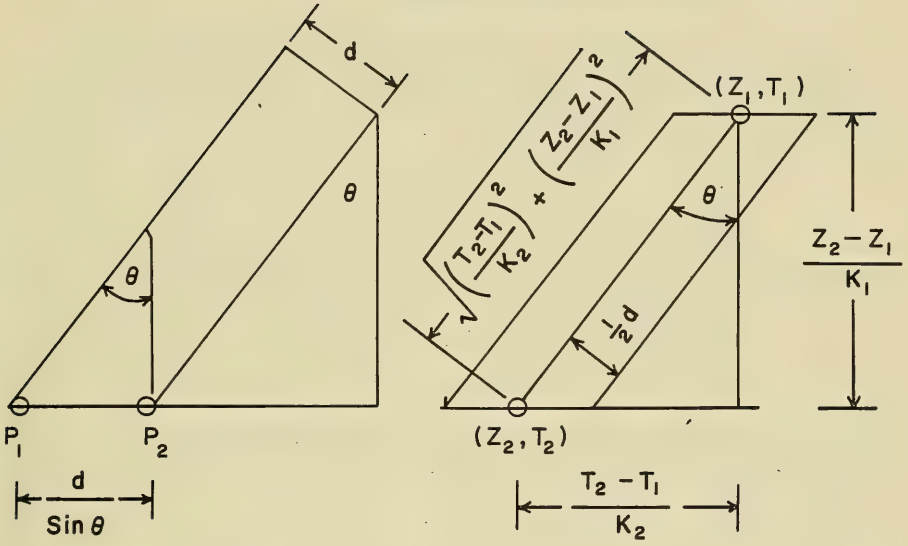


FIGURE 5. (a) THE DISTANCE BETWEEN POINTS P_1 AND P_2 ON A ZONAL SECTION OF THE STYLUS TRACE WILL EXCEED THE DISTANCE, d, BY THE FACTOR 1/sine θ. (b) THE GRAPHICAL GRADIENT ANGLE, θ, IS RELATED TO MEASURES OF TEMPERATURE AND DEPTH (SEE TEXT).

$$e = \frac{\delta T}{2}\sqrt{1 + \left(\frac{k1}{k2} \cdot \frac{T2 - T1}{z2 - z1}\right)^2} \qquad (3)$$

C. Trial and Error Regression Method

It is a simple task to fit a straight line to a point subset by the method of least squares. Assuming the intercept A and B the slope of the line of regression on (m - n + 1) points (z_j, T_j), (j = m, m + 1, . , n), then the solution for A and B is given in matrix notation

$$\begin{bmatrix} A \\ B \end{bmatrix} = (Q^t \cdot Q)^{-1} (Q^t \cdot H),$$

where $Q = \begin{bmatrix} 1 & z_m \\ 1 & z_{m+1} \\ \vdots & \\ 1 & z_n \end{bmatrix}$, $Q^t = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_m & z_{m+1} & \cdots & z_n \end{bmatrix}$, and $H = \begin{bmatrix} T_m \\ T_{m+1} \\ \vdots \\ T_n \end{bmatrix}$ (4)

A set of points may be tested for alignment within the thickness of a stylus trace by trial and error. Given the equation

$$\bar{T}_j = A + B \cdot z_j, \qquad (5)$$

then the points (z_j, T_j) align in the depth interval $z_m \leqq z_j \leqq z_n$, if

$$(T_j - \bar{T}_j)^2 < \left(\frac{\delta T}{2}\right)^2 \cdot \left[1 + \left(\frac{k1}{k2} \cdot B\right)^2\right] \qquad (6)$$

for all values j in the range $m \leqq j \leqq n$.

It is required to identify data point subsets in which as many points as possible align. This may be done likewise by trial and error, using the following method. Proceeding from any point, P, in a set of consecutive profile data points (A, B, C, . .), determine whether points P and Q align. If they do (they always will), then determine whether points P, Q, and R align, and so on. When a set of points is found that does not align, test the previous set for alignment with points preceding P. In this way a point set of maximum alignment extension downward and upward may be identified with any profile data point. Because this set of consecutive points is selected for negligible departure from a straight line, let it be referred to as a "line set".

A minimal number of line sets may be selected initially to cover a profile using the following method. Identify the line set for the surface data point, A, in the profile set (A, B, C, . .). Suppose that the last point in this line set is the point N. Then identify the line set for point N, and so on downward, until the profile data set is exhausted.

Line sets may be said to "overlap" if they contain the same point. They may be called "neighbors" if they contain neighboring points. Line sets that overlap are neighbors, but not all neighboring line sets overlap. The line sets selected to cover a profile by the method described here overlap. That is, for example, the line set defined for point A contains the point, N, as does the line set defined for point N. A line set may have several neighbors, including one that does not overlap. It would be advantageous to represent a profile with neighboring line sets that do not overlap, but this is not always possible. Accordingly, it is required to examine various combinations of line sets for the sake of selecting neighboring sets to represent a profile with minimum overlap.

It is important to distinguish between a line set and the line of regression on the set. The line set, a collection of discrete data points, conveys no information at all about the depth intervals between points. The line of regression is a continuous function without natural limits that articulates implicit assumptions about the intervals between the data points. It must be restricted exclusively to depth intervals containing the line set upon which it regresses. It is therefore convenient to consider a regression line as a line segment which extends to but not across depths of points neighboring the line set (Figure 6(a)). So considered, it is meaningful to refer to "neighboring" regression lines. It may be noted that regression lines on neighboring line sets overlap the depth interval between the two sets whether or not the line sets overlap. Hence, it is meaningful to refer to this depth interval as the "range of overlap" between neighboring regression lines.

Suppose that neighboring line sets respectively contain points (A, B, C) and (D, E, F). The range of overlap between regression lines on those two sets is the depth interval between points C and D. The right limit of the range of overlap may be identified as the depth of the point which neighbors the left set on the right, and the left limit may be identified as the depth of the point which neighbors the right set on the left. When there is overlap between neighboring sets the range of overlap will contain one or more data points. Thus point E would be the right limit and point B the left limit of the range of overlap between regression lines on sets (A, B, C, D) and (C, D, E, F), and the range would contain points C and D.

D. Intersection

Neighboring regression lines may or may not intersect in the range of overlap. If they intersect, then the depth of intersection may be used to represent the limit of regression lines on neighboring line sets. The two lines will form a path, or "union", that will be within half a stylus trace thickness of data points contained in the two line sets. If they do not intersect in the range of overlap, alternative methods must be used to represent the union of neighboring line sets. The path derived by extending the regression lines to the point of intersection outside the range of overlap will fail to represent all data points in the union of neighboring sets with predictable precision (Figure 6).

Neighboring regression lines frequently do not intersect in the range of overlap when their line sets do not overlap. Hence it is necessary to examine alternative combinations of line sets that overlap when searching for a union of regression lines to represent a profile. Thus, for example, when the regression lines on the line sets (A, B, C) and (D, E, F) do not intersect between points C and D, then an intermediate line set, defined for point C, must be considered. Occasionally, when they are nearly parallel, the regression lines for overlapping line sets will not intersect in the range of overlap. In this case the regression lines must be close together in the range of overlap, since both lines must fall within half a stylus line thickness of data points contained in the range.

The following logic therefore seems to be suggested as a means for selecting a union of lines to fit a digital profile: (1) Starting from the surface, identify a "base" regression line, and select a neighboring regression line to represent a union with the base line; (2) When a neighbor is selected, identify the neighbor as a base line, and repeat the selection process downward, until the set of regression lines is exhausted. Neighboring regression lines may be examined and selected for the behavior of intersection with the base line according to the following scheme: (1) If the point of intersection falls in the range of overlap, tentatively identify it as a "regression point"; (2) If the point of intersection does not fall in the range of overlap, determine whether their line sets overlap. If they do not overlap, then the union of the base regression line with its overlapping neighbor must be examined, repeating the procedure above. If they do overlap then the median depth in the range of overlap and the median temperature between the two regression lines at that depth may tentatively be selected as a substitute regression point between the two line sets. This point is within half a stylus trace thickness of the regression lines on the two line sets, and constitutes a tolerable compromise.

A set of regression points may thus be tentatively defined to represent a union of straight lines fitted to a profile data set. The union may be regarded

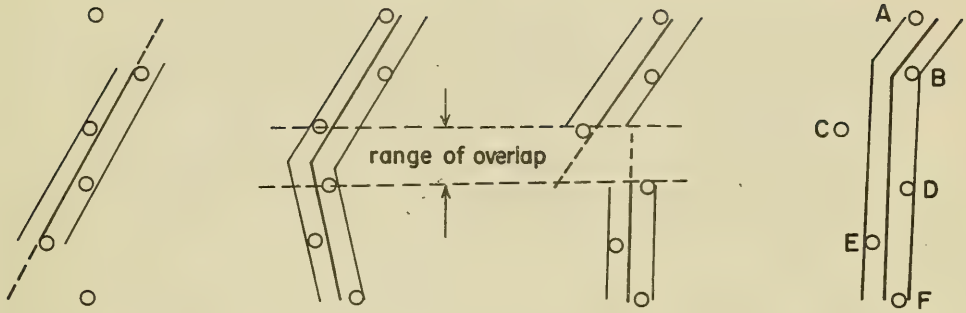


FIGURE 6. (a) THE REGRESSION LINE IS A "LINE SEGMENT" THAT EXTENDS OVER THE DEPTH INTERVAL OF POINTS IN ITS LINE SET, BUT NOT OVER THE DEPTHS OF NEIGHBORING POINTS. (b) NEIGHBORING LINE SEGMENTS MEET IN THE "RANGE OF OVERLAP" TO FORM A "UNION." THE POINT AT WHICH THEY MEET IS CALLED A "REGRESSION POINT." (c) NEIGHBORING LINE SEGMENTS DO NOT MEET IN THE RANGE OF OVERLAP. IF THE LINE SEGMENTS ARE EXTENDED TO THE POINT OF INTERSECTION (d) THEN AT LEAST ONE DATA POINT (IN THIS INSTANCE, POINT C) WILL NOT BE WITHIN HALF A STYLUS THICKNESS OF THE RESULTING UNION.

as a regression function discontinuous in the first derivative at the points of intersection between neighboring lines. Some of these points may subsequently be deleted from the set of regression points finally selected to represent the profile.

E. Smoothing

It happens that the tentative set of regression points may contain virtually colinear points. That is, occasionally a point in the set may be a negligible distance from a straight line connecting two points on alternate sides of it. This point may be deleted and a single regression line may be computed to fit the combined line sets on either side of it, provided the resulting union of regression lines does not deviate by an intolerable distance from any data point. It would therefore seem advisable to provide definitions of a "negligible distance" between a regression point and a straight line, and an "intolerable distance" between a data point and a regression line, as criterion for smoothing a set of regression points.

Accordingly, a "three level tolerance algorithm" (Figure 7) has been devised for reducing a profile data set to a select set of regression points. The first level of tolerance is used to define a "thin line", for assembling an initial set of thin line sets to fit the profile data. The second level provides the definition of a negligible distance between a regression point and a straight line. The third level provides the definition of intolerable distance between a data point and a regression line. A computer program (appendix) was developed to analyze a profile data set and smooth the resulting set of regression points by the trial and error methods outlined here, using the somewhat arbitrary ratios $1:\sqrt{2}:2$ to represent the three levels of tolerance. The union of regression lines derived with this program does not deviate from the digital data profile by more than the thickness of a "thin" stylus line ($\pm 2e$).

F. Width of the "Thin" Stylus Line

Initial estimates provided by actual BT trace measurements for stylus trace thickness were refined according to the following scheme.

Several BT profiles were visually digitized to serve as data for the development of the profile analysis algorithm and program. At a later time, without reference to the initial data aggregates, the same profiles were visually digitized a second time. Differences between temperatures at corresponding depths in corresponding aggregates were calculated. Individual temperature differences were corrected for trace slope to provide an estimate of the component of difference normal to the stylus trace in the solution plane.

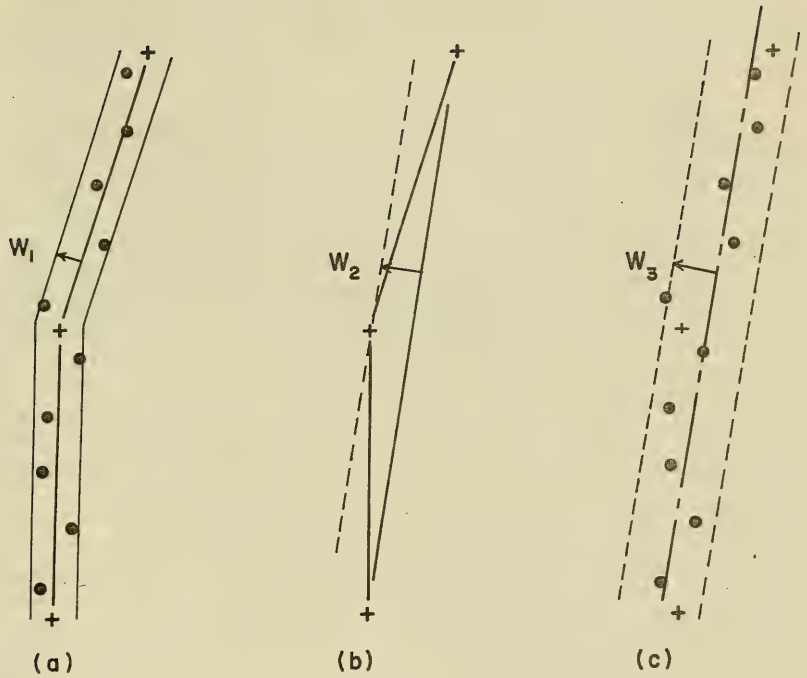


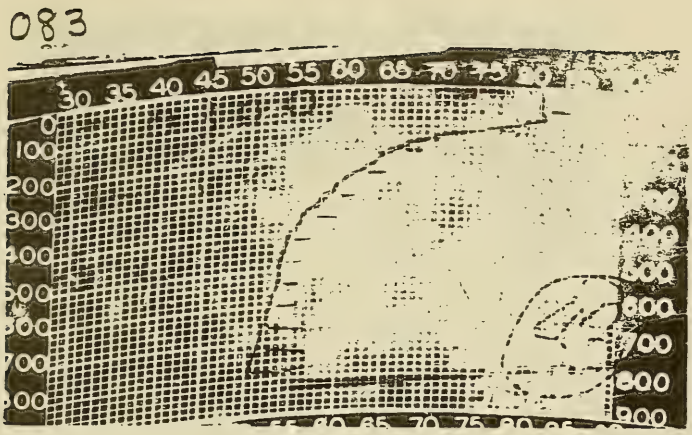
FIGURE 7. THE "THREE LEVEL TOLERANCE" ALGORITHM USES A FIRST LEVEL TOLERANCE VALUE, W_1, TO DETECT "THIN LINE SETS" IN A DIGITAL PROFILE (a). A SECOND LEVEL TOLERANCE VALUE, W_2, IS USED TO DETECT "VIRTUALLY COLINEAR" POINTS IN A REGRESSION SET (b). WHEN A VIRTUALLY COLINEAR POINT IS DETECTED A NEW REGRESSION SET IS COMPUTED, COMBINING LINE SETS ON EACH SIDE OF THE COLINEAR POINT. A THIRD TOLERANCE LEVEL VALUE, W_3, IS USED TO DETERMINE WHETHER THERE IS A DATA POINT THAT IS AN "INTOLERABLE" DISTANCE FROM THE RESULTING REGRESSION SET (c).

The square root of the mean of the squares of corrected differences was then calculated to provide an index of comparison for corresponding data aggregates. The root mean square values obtained ranged from about .09° to .15° Fahrenheit for the samples analyzed.

Corresponding data aggregates were then computer processed for regression point sets using different values for δT. An index of comparison was computed for corresponding regression sets, based upon differences between interpolated temperatures at corresponding five meter intervals corrected for trace slope. It was reasoned that if it is appropriate to fit straight line segments to a digital BT, then for some values of δT the index of comparison would be less than the index of comparison for the original data aggregates. Moreover, the index of comparison would tend to diminish to a minimum for values of δT representing the thickness of the stylus trace. It was found that for values of δT in the neighborhood 0.2° Fahrenheit, the differences between the regression sets were somewhat less than the differences between the original data aggregates. For values of δT in the neighborhood 0,3° Fahrenheit, the index of comparison for corresponding regression sets ranged in value for about 0.06°, a minimum for the data samples and δT values tested, to .15° Fahrenheit. The most conclusive results were obtained for profiles with smooth gradients and slight hysteresis (Figure 8) and the least conclusive results were obtained for irregular profiles (Figure 9). On the basis of these experiments it was concluded that the most appropriate choice for δT was probably about 0.3° Fahrenheit (0.16° Celsius), or about .015 inches on the plane of the standard scale bathythermograph.

G. Measure of "Appropriateness of Fit"

It was considered advisable to provide a measure of the appropriateness of fit of a regression line over the interval of regression. Accordingly, a subroutine was used in the regression program to calculate the root mean of the squares of differences between temperature data in the interval between regression points and interpolated temperature values at corresponding depths on the regression line. These values were not corrected for the slope of the line. It was assumed that the user of the BT data would be more interested in estimating the reliability of temperatures in a profile than the probable departure of a point from a straight line on the bathythermogram. It follows that the RMS values for the profile interval between adjacent regression points will tend to increase with increasing slope angles and decrease to a minimum for a vertical trace (slope angle zero). If the user is interested in relating these RMS values to departures from a straight line in the plane of the bathythermogram he may use the relationship:



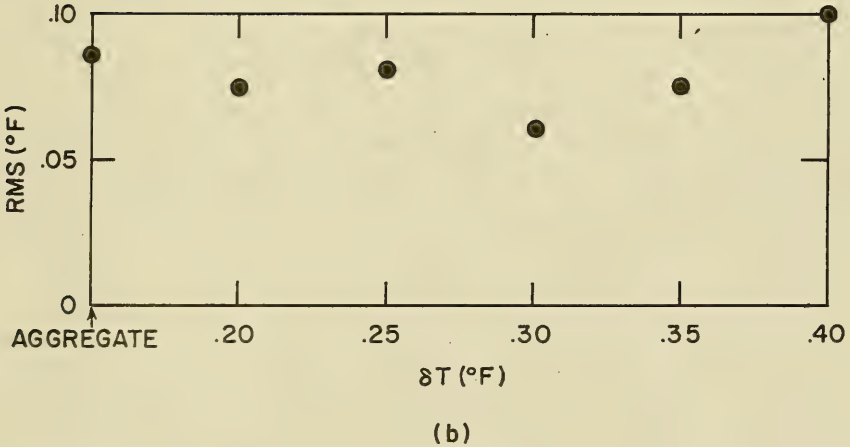
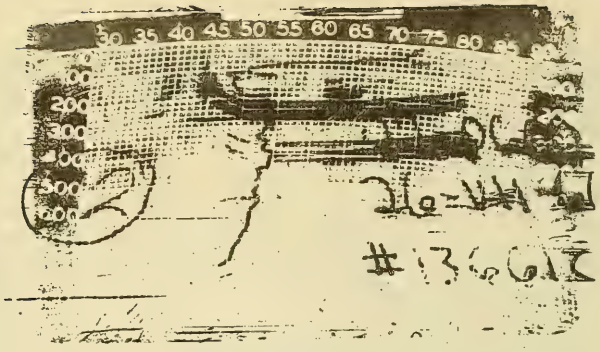
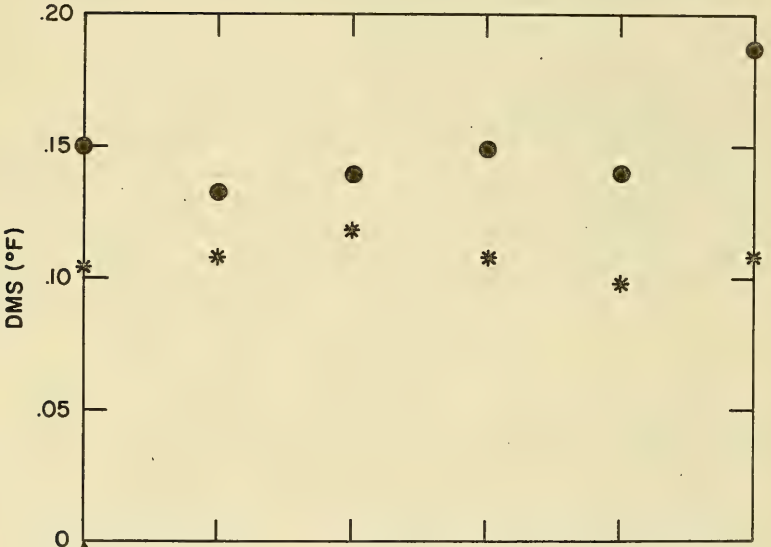


FIGURE 8. THE BT PROFILE (a) WAS DIGITIZED AT TWO DIFFERENT TIMES. THE DATA AGGREGATES WERE PROCESSED FOR LINES OF FIT WITH DIFFERENT δT VALUES AND THE RESULTS WERE COMPARED (b). THE MINIMUM DEPARTURE BETWEEN LINE UNIONS WAS FOUND FOR THE VALUE $\delta T = .3°$ FAHRENHEIT. IT WAS CONCLUDED THAT THE MOST APPROPRIATE "STYLUS THICKNESS" VALUE FOR THIS PROFILE IS ABOUT .3° FAHRENHEIT OR .16° CELSIUS.



(b)



(a)

FIGURE 9. THE BT PROFILE (a) WAS DIGITIZED TWICE. NOTE PROFILE IRREGULARITY AND POOR READABILITY. THE TWO DATA AGGREGATES WERE PROCESSED FOR LINES OF FIT WITH DIFFERENT VALUES FOR δT AND THE RESULTS WERE COMPARED (b – dots). THE RESULTS WERE COMPARED EXCLUDING VALUES IN THE INTERVAL OF ERASURE (b – ASTERISKS). THE RESULTS SUGGESTED INCONCLUSIVELY THAT A VALUE FOR δT OF ABOUT .35° WAS MOST APPROPRIATE FOR THIS PROFILE.

$$LRMS = RMS/\sqrt{1 + (k \cdot B)^2}, \qquad (7)$$

where RMS is the recorded root mean square of departures between temperatures on a regression line and the temperature data that the line is supposed to fit; LRMS is the RMS corrected to represent departure of data points from a regression line in the bathythermogram, in units of temperature; k is the constant relating measures of depth and temperature (equals about 14.5 meters per degree Celsius); and B is equal to $\frac{T_2 - T_1}{z_2 - z_1}$, where (z_1, T_1) and (z_2, T_2) are the regression points bracketing the interval of interest.

The RMS is provided as a measure of the appropriateness of fit of a regression line to a set of data points. It may be interpreted as an indication of profile roughness over the range of the line, as an index of random scatter of data points within the thickness of a smooth stylus trace, or as a combination of these factors. If there is ambiguity in the choice of interpretations available there is likewise ambiguity in the interpretation of digitized points that provide the lines of fit.

COMMENTS AND CONCLUSION

A. Room for Analytical Studies

The profile regression problem and the method of solution described here provide a rich source for extensive statistical studies, with inquiries concerning: comparisons between the graphical method of solution using relationships of measure taken from the standard scale bathythermograph and alternative solutions; the appropriateness of fitting straight lines to a profile data set, as opposed to fitting other types of curves; alternatives to selecting a median regression point when regression lines on overlapping line sets do not intersect in the range of overlap; possible variations of the three level tolerance algorithm using different tolerance ratios, and comparisons; the application of this solution algorithm to profile data sets other than those afforded by mechanical BT's; the best method for selecting regression filters, or "line thicknesses", to process mass quantities of profile data.

B. The "Three Level Tolerance Algorithm"

The three level tolerance algorithm was devised to simulate a graphical approach to detecting profile line sets with ruler and pencil. That is, a graphical approach to digitizing significant depths on a BT would be: (1) identify profile segments that obviously constitute straight lines (identify an initial set of thin line sets); (2) identify points of intersection between neighboring straight

line segments; (3) delete short amplitude irregularities from the set of intersection points, provided resulting deviations from the profile are not excessive.

Experimental evidence suggests that, in general, a large "thin line" tolerance tends to smooth the data set too much, yielding line sets with excessive overlap. On the other hand, a small value for the "negligible" departure between a regression point and a straight line tends to yield unions of lines consisting of more line segments than are necessary, and excessive values for the "intolerable" departure between a straight line and a data point results, in some instances, in an unacceptable loss of profile regression precision. By varying the ratio of these tolerances it was determined that the ratio finally selected provides advantages of minimum data smoothing (small RMS departures) with gratifying economy in the number of lines selected to represent a profile.

C. Compromise Regression Points

The following method is used to calculate a compromise regression point when regression lines on overlapping line sets do not intersect in the range of overlap. If the depths z_1 and z_2 are the limits of the range of overlap, then an estimate for the regression depth is provided:

$$\bar{z} = (z_1 + z_2)/2. \tag{8}$$

If the equations of the regression lines are provided, respectively,

$$T = A_1 + B_1 \cdot z \tag{9}$$

$$T = A_2 + B_2 \cdot z \tag{10}$$

then the median point in the range of overlap is identified by the coordinates $(\bar{z}, \bar{T})$, where

$$\bar{T} = \left[(A_1 + A_2) + (B_1 + B_2)\, \bar{z} \right] /2 \tag{11}$$

It may be noted that this method will not under any circumstance yield an error exceeding a thin stylus line departure ($\pm 2e$) between the regression profile and any profile data point. Hence, this method is consistent with the three level tolerance algorithm that permits the same maximum departure.

D. Fitting Curvelinear Functions

The three level tolerance algorithm for fitting lines to points in a plane may be applied with appropriate modifications to the fitting of alternative sets

of curvelinear functions to points in a plane. The advantage of alternative functions would be that they could be used to estimate second and perhaps higher derivatives of profile distribution as well as profile gradients.

It may be pointed out that the straight line is the least expensive function that can be fit to a set of points in a profile, in terms of the computer time required to identify a union of regression curves on a profile data set and perhaps in terms of the computer memory required to record the union as well. A profile may be represented economically with N + 1 regression points for a union of N regression lines, whereas more complicated methods would have to be used to represent a union of N higher order regression curves. Moreover, a solution has not been specified here to the problem of representing the union of neighboring curve sets that do not intersect in the range of overlap, or which intersect there more than once (Figure 10).

The straight line probably constitutes a function of sufficient resolution to represent standard scale mechanical BT profiles. It is doubtful that higher order functions can be used advantageously to gain either precision or economy of BT data representation. This is not to say that the straight line is the most appropriate function for representing other kinds of profile data sets, however.

E. Other Kinds of Profile Data Sets

When applying this algorithm to the reduction of a profile data set it is necessary to select a coordinate system that will appropriately relate measures of profile interval to measures of profile amplitude. The choice of the bathythermograph slide scale seems to be an obvious one in the case of the standard mechanical BT. In general, the selection must be rendered on the basis of less obvious criterion.

As is the case for the mechanical BT, gradient features of a profile, rather than individual trace values, may be worth preserving. In order to record these features with reasonable fidelity it may be necessary to record individual trace values to tolerances that would seem unrealistic considering instrument inaccuracy. Evidently, when one sensor is used to measure a profile, instrument error over the profile range may be regarded as systematic. That is, the sensor tends to stay inaccurate by the same amount over short time intervals during the measurement of the profile. These inaccuracies may accumulate over an extended time period or with changing circumstances, as is evidenced by the hysteresis in a typical BT profile trace.

Accordingly, it may be suggested that the coordinate scale be designed to reflect the sensitivity, rather than the absolute accuracy, of the profile

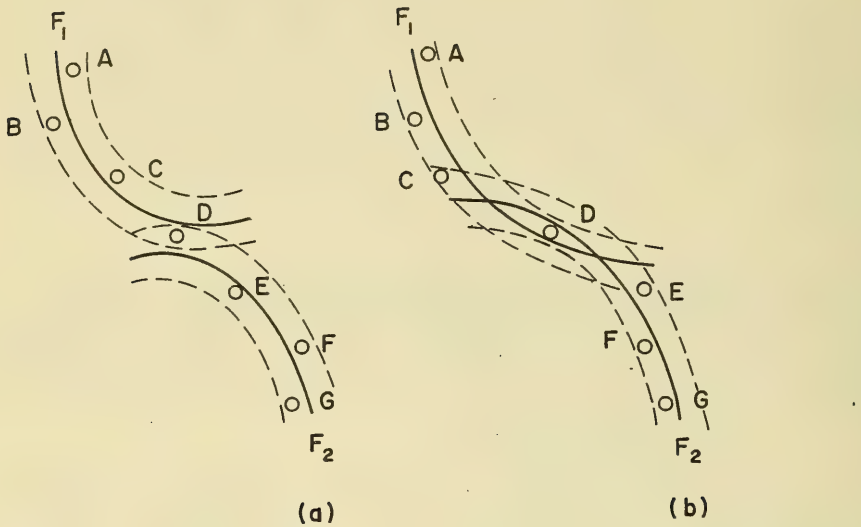


FIGURE 10. FITTING CURVELINEAR FUNCTIONS: (a) F_1 IS A REGRESSION FUNCTION ON THE FUNCTION SET (A,B,C,D) AND F_2 IS A REGRESSION FUNCTION ON THE FUNCTION SET (D,E,F,G). THE TWO FUNCTIONS DO NOT INTERSECT IN THE RANGE OF OVERLAP; (b) F_1 IS A REGRESSION FUNCTION ON THE FUNCTION SET (A,B,C,D,E) AND F_2 IS A REGRESSION FUNCTION ON THE FUNCTION SET (C,D,E,F,G). THE TWO FUNCTIONS INTERSECT TWICE IN THE RANGE OF OVERLAP.

sensor. That is, if the threshold of sensitivity (the smallest change to which a sensor will respond with a significant measure of reliability in a meaningful time interval) is "A" units of measure in profile amplitude and "B" units of measure in profile interval, then it may be suggested that this threshold interval serve as an initial estimate of the line thickness in the graphical solution plane, and that the grid of the plane be chosen such that A units of amplitude equate to B units of interval.

PROFILE REGRESSION PROGRAM

Subroutine "REDUCE" (Appendix) may be used to fit sets of straight lines to digital BT profiles in the National Oceanographic Data Center format. This program has been used in conjunction with a master program for restructuring the NODC BT file for quick archival and retrieval.

A. Input/Output Data Linkage

Data linkage with the users main program is achieved with the use of a common memory block that is labeled "PROFILE". The number of data points, N, and the digital profile, T(J), (J=1, N) are specified as input data to the subprogram. It is assumed that the digital profile is represented in Celsius temperature values at five meter intervals, starting from depth zero for J=1.

The number of profile regression points, M, regression depths, DL(L), (L=1, M), regression temperatures, TL(L), (L=1, M), and root mean square values of differences between data and corresponding regression temperatures in the intervals of fit, RMS(L-1), (L=2, M), are specified as output from this program. Temperatures and RMS values are Celsius degrees. Depths are expressed in the scale of input subscript intervals, and may be converted to meters with the equation

$$D(J) = 5.*(DL(J) - 1.0). \tag{12}$$

The root mean square (RMS) values are computed for intervals, commencing with the interval between DL(1) and DL(2), and ending with the (M-1)th interval, between DL(M-1) and DL(M). The arguments summed are the differences

$$(T(J) - \bar{T}_i)^2 \tag{13}$$

where $\bar{T}_i = A_k + B_k \cdot J$

$$A_k = \frac{TL(k-1) \cdot DL(k)}{DL(k) - DL(k-1)}$$

$$B_k = \frac{TL(k) - TL(k-1)}{DL(k) - DL(k-1)}$$

for values of J in the range $DL(k-1) \leqslant J \leqslant DL(k)$.

B. Subroutines

Routines that are used repeatedly are coded as subroutines. These include subroutines "TF", which relates temperature and depth intervals to measures on the scale of the bathythermogram, "SQLINE", which calculates the equation of the regression line on a set of points using the method of least squares, and "CORD", which calculates by trial and error the limits of a thin line set identified with a base data point. Subroutine "TF" is coded in Compass, the assembly language for the CDC-3600 computer series. It may be replaced, if desired, by the fortran function statement:

$$TF(B) = .0064 + .0565*B**2 \qquad (14)$$

where B is temperature slope expressed in Celsius degrees per subscript unit interval of depth.

The constants in the function TF may be modified to represent different graph scales and tolerances. Thus, by modifying this function appropriately, subprogram "REDUCE" may be used to reduce digital profile data from different sources on different graph scales, provided the data are digitized at regular intervals. If the data are not digitized at regular intervals other portions of the subprogram, including "SQLINE", would have to be modified.

C. Reduction of Data Memory Volume Requirements

The NODC global BT file (circa June 1968) was processed with this subprogram, and the regression profiles were recorded in a "restructured" BT file. Preliminary studies indicate that the thin line tolerance value ±0.08° Celsius, permits a volume reduction of about four or five data points to one regression point. A more precise estimate, including an analysis of variation with depth and geographical area, may be determined by an exhaustive survey of the restructured file.

Depths of data points were not indicated in the original NODC BT profile format, save by data position relative to the first data value reported. The depth of a regression point must be more explicitly identified. Hence, some portion of the savings effected in the number of points required to represent a profile is absorbed by the weight of additional data needed to specify the depths of regression points. If it is assumed that an equal number of digits are needed to encode depth and temperature, then it follows that the volume of the BT profile data has been reduced by a factor of 2 or 2.5 with this program. This factor is further reduced if the RMS values are considered as part of the product volume of the program.

D. Production Time

On an average, the program reduces about seven profiles per second or twenty five thousand profiles per hour on the CDC 3800 computer facility at the Naval Research Laboratory. This figure varies with the depth and distribution regularity of a profile. It is estimated that the average profile which provides this figure consists of about forty data points, and reduces to about ten regression points.

The major portion of program production time is consumed in the selecting by trial and error of the initial set of thin line sets to fit the profile data. This time may be reduced significantly by optimizing the subroutines "TF", "SQLINE", and "CORD", and recoding them in assembly language.

APPENDIX A

SUBROUTINE REDUCE	A-1
SUBROUTINE SQLINE	A-4
SUBROUTINE CORD	A-5
SUBROUTINE TF	A-6
Sample BT Listing	A-7

DISCLAIMER

The U. S. Naval Oceanographic Office will not be liable for the performance of this program in any context of private, civil or industrial use. The user is responsible for adapting this program to his facility if he chooses, and he is accordingly responsible for the program success or failure.

The author expresses a personal interest in the performance of this program, however, and would appreciate evidence that may be used to detect and correct any unforeseen program anomalies.

```
      SUBROUTINE REDUCE
      COMMON/PROFILE/N,T(90),M,DL(90),TL(90),RMS(90)
      DIMENSION A(90),B(90),MIN(90),MAX(90),IPOINT(90)
C
C  ***************************************************************
C     SUBPROGRAM FOR CALCULATING LINES OF FIT FOR
C                  NODC FORMATTED DATA
C         YERGEN     R+E     NAVOCEANO     5/23/68
C  ***************************************************************
C
      IPOINT(1)=1  $ DL(1)=1.
C
C*****************************************************************
C
C   IDENTIFY AN INITIAL SET OF THIN LINE SETS (CORDS) IN THE DATA.
C
C*****************************************************************
C
        L=0  $ I=2
   12 CALL CORD(T(1),J,N,NUP,NDOW,A1,B1)
      IF(L) 14,14,15
   15 DO 18 K=1,L
      IF(NUP-MIN(K))  13,16,18
   16 IF(NDOW-MAX(K))  13,13,17
   17 MAX(K)=NDOW  $ A(K)=A1  $ B(K)=B1
        IF(NDOW-J)  13,13,21
   13 J=J+1  $ IF(N-J)  22,12,12
   18 CONTINUE
   14 L=L+1  $ MIN(L)=NUP$MAX(L)=NDOW$ A(L)=A1 $ B(L)=B1
   21 J=NDOW    $IF(N-J)  22,22,12
C
C     ***************************************************
C
C   SELECT A SET OF NEIGHBORING CORDS, MINIMIZING CORD OVERLAP.
C
C     ***************************************************
C
   22 M=1
   23 K=IPOINT(M) $ ITRY=MAX(K)+1
      K2=K+1
   24 K2=K2+1  $ IF(L-K2) 26,25,25
   25 IF(ITRY-MIN(K2))  26,24,24
   26  K2=K2-1  $  IF(K2-K)  259,259,260
  259  CONTINUE
        K2=K2+1  $  IF( L-K2)  31,261,261
  261  RLIM=MAX(K)+1
       HLIM=MIN(K2)-1 $  M=M+1 $ IPOINT(M)=K2
       DL(M)=.5*(HLIM+RLIM)
       TL(M)=.5*(A(K)+A(K2)+(B(K)+B(K2))*DL(M))
         IF(N-MAX(K2))  31,31,23
  260 RLIM=MAX(K)+1  $ HLIM=MIN(K2)-1   $ DI=B(K2)-B(K)
      IF(DI)  28,26,28
   28 DI=(A(K)-A(K2))/DI
      IF(RLIM-DI)  26,29,29
   29 IF(DI-HLIM)  26,30,30
   30 M=M+1  $ IPOINT(M)=K2 $ DL(M)=DI $ TL(M)=A(K)+DI*B(K)
```

```
      IF(N-MAX(K2))  31,31,23
   31 M=M+1  $ DL(M)  =N
      TL(M)=A(L)+B(L)*DL(M)
C
C     ***************************************************
C
C     ****  RECURSIVE SMOOTHING FUNCTION  ****
C   DELETE REGRESSION POINTS THAT FALL WITHIN A NEGLIGIBLE DISTANCE
C   OF A STRAIGHT LINE CONNECTING ALTERNATE REGRESSION POINTS.
C
C     STEP 1   TEST SUCCESSIVE POINT SETS FOR COLINEARITY
C
C     ***************************************************
C
  301 IVY=0  $ L=0  $ J=1
  302 K=J+2  $ IF(M-K) 307,303,303
  303 B3=(TL(K)-TL(J))/(DL(K)-DL(J))
       A3=TL(J)-B3*DL(J)
       Q= (A3+B3*DL(J+1)-TL(J+1))**2
      PRM=1.414213562*TF(B3)
  308 L=L+1  $ KI=IPOINT(J)
      MIN(L)=MIN(KI)  $ IPOINT(L)=L
      IF(PRM-Q)  305,304,304
C
C     DELETE SUPERFLUOUS INTERMEDIATE POINTS
C
  304 CONTINUE
      NUP=MIN(KI)$KU=IPOINT(J+1)$NDOW=MAX(KU)
      CALL SQLINE(T(1),NUP,NDOW,A1,B1)
      PRM=2.*TF(B1)
      DO 32 NN=NUP,NDOW
      Q=NN  $ Q=(A1+B1*Q-T(NN))**2
      IF(PRM-Q)  305,305,32
   32 CONTINUE
      IVY=1  $ KI=KU
      A(L)=A1  $ B(L)=B1  $ J=J+1$ GO TO 306
  305  A(L)=A(KI)  $ B(L)= B(KI)
  306 MAX(L)=MAX(KI)  $ J=J+1$  IF(M-J)  310,310,302
  307  Q=PRM+10.
       GO TO 308
  310 CONTINUE
C
C     STEP 2   IDENTIFY REGRESSION POINTS
C
      M=1  $ TL(1)=A(1)+B(1)
  395  IF(L-M)  45,45,40
   40 IM=M+1$RLIM=MAX(M)+1$HLIM=MIN(IM)-1
      DI=(A(IM)-A(M))/(B(M)-B(IM))
      IF(RLIM-DI)  41,42,42
   41 DL(IM)=.5*(RLIM+HLIM)
      TL(IM)=.5*(A(M)+A(IM)+(B(M)+B(IM))*DL(IM))
      GO TO 44
   42 IF(DI-HLIM)  41,43,43
   43 DL(IM)=DI
      TL(IM)=A(M)+B(M)*DI
   44 M=IM$GO TO 395
```

```
   45 M=M+1  $ DL(M)=N  $ TL(M)=A(L)+B(L)*DL(M)
      IF(IVY)  50,50,301
   50 CONTINUE
C
C
C  ***************************************************************
C
C     CALCULATE RMS DEPARTURE FROM ORIGINAL TEMPERATURE DATA
C
C       NOT CORRECTED FOR SLOPE
C
C  ***************************************************************
C
       DO 409 K=2,M
       RSM=0   $ RCT=0
       X1=DL(K-1)  $ X2=DL(K)$  A1=A(K-1) $  B1=B(K-1)
        DO 405  I=1,N
      C=I  $  IF(X1-C)  403,403,405
  403  IF(C-X2)  404,404,406
  404  TJ=A1+C*B1
      RSM=RSM+(T(I)-TJ)**2
       RCT=RCT+1.
  405  CONTINUE
  406  IF(RCT)  408,408,407
  407  RCT=SQRTF(RSM/RCT)
  408  RMS(K-1)=RCT
  409  CONTINUE
       RETURN
         END
```

```
      SUBROUTINE  SCLINE(T   ,K1,K2,A,B)
C
C ***************************************************************
C     CALCULATES LEAST SQUARES LINE OF REGRESSION FOR DT DATA POINTS
C                FROM T(K1) TO T(K2), INCLUSIVE
C       LINE    T=A+B*J   FOR  J=K1,...,K2
C
C     ***************************************************
      DIMENSION  T(1)
      A=0.$B=0.$SD=0.$SD2=0.$ST=0.$STD=0.
      DO 5 J=K1,K2
      DT=J-K1 $ TJ=T(J)$SD=SD+DT$SD2=SD2+DT**2
      ST=ST+TJ$STD=STD+DT*TJ
    5 CONTINUE
      DT=K2-K1+1$SD=SD/DT$SD2=SD2/DT$ST=ST/DT$STD=STD/DT
      DT=1./(SD2-SD*SD)$SD2=SD2*DT
      SD=-SD*DT$A1=SD2*ST+SD*STD$B1=SD*ST+DT*STD
      DT=K1$A=A1-B1*DT$B=B1 $ RETURN
         END
```

```
      SUBROUTINE CORD(T,J,N,JM,M,A,B)
      DIMENSION  T(1)
C
C     ************************************************************
C     IDENTIFIES MINIMUM AND MAXIMUM SUBSCRIPTS, JM AND M,
C     OF POINTS T(JM) THROUGH T(M) IN A CORRIDOR OF LEAST SQUARE
C     DIFFERENCES EXTENDED FIRST DOWNWARD AND THEN UPWARD FROM POINT
C     T(J).  LINE OF LEAST SQUARES IS GIVEN BY THE EQUATION
C              T=A+B*K FOR  K=JM,M
C     MAXIMUM ERROR (T(K)-T)**2 IS EQUAL TO OR LESS THAN TF(B)
C     VERGEN    NAVOCEANO     CODE 7240
C     ************************************************************
C
      JM=J-1  $  M=JM $IZIP=1
    1 M=M+1$IF(N-M)  5,2,2
    2 CALL SCLINE(T(1),JM,M,A1,B1)  $ PRM=TF(B1)
      DO 3 L=JM,M
      Q=L$Q=(A1+B1*Q-T(L))**2
      IF(PRM-Q)  7,3,3
    3 CONTINUE  $  A=A1  $ B=B1
      IF(IZIP)  4,1,1
    4 JM=JM-1  $ IF(JM) 6,6,2
    5 M=M-1  $IZIP=-1$  GO TO 4
    6 JM=JM+1 $ RETURN
    7 IF(IZIP)  6,5,5
      END
```

```
                                                IDENT  TF
PROGRAM LENGTH                    00031
ENTRY POINTS     TF               00013
                 TFSCALE          00000

                                                ENTRY  TF
                                                ENTRY  TFSCALE
00000    00  0  00000   TFSCALE   OCT
         00  0  00000
00001    77  1  04000             ENO   *
         57  1  P00011             SIL   TFSX,1
00002    12  0  P00000             LDA   TFSCALE
         00  7  40541             RXT   A,B1
00003    01  0  00017             ARS   15
         00  7  40565             RXT   A,@B
00004    12  1  00001             LDA   1,1
         00  7  40401             RXT   AU,B1
00005    01  0  00047             ARS   39
         00  7  40565             RXT   A,@B
00006    12  1  00000             LDA   ,1
         77  1  04000             ENO   *
00007    20  0  P00026             STA   =SA
         32  0  P00026             FMU   A
00010    20  0  P00026             STA   A
         32  0  P00025             FMU   =DA,8209
00011    20  0  P00027   TFSX      STA   =SB
         50  1  00000             ENI   ,1
00012    75  0  P00000             SLJ   TFSCALE
         50  0  00000
00013    00  0  00000   TF        OCT
         00  0  00000
00014    77  1  04000             ENO   *
         56  1  P00024             SIU   TFX,1
00015    12  0  P00013             LDA   TF
         00  7  40541             RXT   A,B1
00016    01  0  00017             ARS   15
         00  7  40565             RXT   A,@B
00017    12  1  00001             LDA   1,1
         00  7  40401             RXT   AU,B1
00020    01  0  00047             ARS   39
         00  7  40565             RXT   A,@B
00021    12  1  00000             LDA   ,1
         77  1  04000             ENO   *
00022    20  0  P00030             STA   =SX
         32  0  P00030             FMU   X
00023    32  0  P00027             FMU   B
         30  0  P00026             FAD   A
00024    50  1  00000   TFX       ENI   ,1
         75  0  P00013             SLJ   TF
00025    20  0  44322
         11  5  00235

                                                END
```

BT PROFILES NODC DATA * YERGENS FORMAT *

MSQR DEG M0

042 47 APR 19 1961 TIME 0300
REF 04156 14 37 N 057 59 W
CSC 887 QUAL 2
BOTTOM DEPTH 4572 M
WIND 18 KNOTS, E
WAVES 3
TD=26.7 TW=23.3
P=1015 MB
CUMULUS COVER 2

	DEPTH M	TEMP C	GRAD C/M	RMS C
1	0	26.9		
2	19	26.7	-0.01	0.06
3	43	26.3	-0.02	0.06
4	81	26.1	-0.00	0.06
5	97	26.4	0.02	0.06
6	118	24.9	-0.07	0.13
7	128	23.9	-0.09	0.13
8	131	23.8	-0.06	0.06
9	166	20.9	-0.08	0.13
10	177	19.8	-0.10	0.06
11	220	17.6	-0.05	0.06
12	240	17.1	-0.03	0.06

042 47 APR 30 1966 TIME 2000
REF 07126 14 09 N 057 53 W
CSC 335 QUAL 2
BOTTOM DEPTH 4206 M
WIND 3 KNOTS, NE
WAVES 1
CUMULUS COVER 2

	DEPTH M	TEMP C	GRAD C/M	RMS C
1	0	26.8		
2	27	26.7	-0.00	0.06
3	43	25.9	-0.05	0.06
4	95	24.9	-0.02	0.06
5	137	20.4	-0.11	0.06
6	158	18.9	-0.07	0.13
7	167	18.5	-0.05	0.06
8	210	17.1	-0.03	0.06

042 47 MAY 18 1959 TIME 2005
REF 06132 14 16 N 057 19 W
CSC 499 QUAL 2
BOTTOM DEPTH 5029 M

	DEPTH M	TEMP C	GRAD C/M	RMS C
1	0	27.3		
2	36	26.8	-0.01	0.06
3	71	25.7	-0.03	0.06
4	84	25.5	-0.01	0.06
5	100	24.9	-0.04	0.06
6	121	23.5	-0.07	0.06
7	162	18.7	-0.12	0.13
8	180	18.4	-0.02	0.06
9	185	17.4	-0.20	0.00
10	215	16.0	-0.05	0.06

APPENDIX: Sample listing of BT profiles, retrieved from the restructured file. Regression points are listed, along with the gradients and RMS departures for successive intervals between regression points.

U.S. Naval Oceanographic Office
FITTING A SET OF STRAIGHT LINES TO A DIGITAL BT PROFILE, by W.E. Yergen, April 1969, 31 p. including figures and appendix. (TR-214)

A computer algorithm for fitting straight lines to subsets of points in a digital BT profile is described. A three level tolerance algorithm for selecting subsets of points that align is submitted, along with a method for determining an optimum thin line tolerance value. The algorithm may be applied to any kind of digital profile data set.

1. Algorithm, three level tolerance for fitting straight lines to point subsets in a digital profile.
2. Bathythermograph data reduction and representation.
3. Digital profile analysis algorithm.

i. title: Fitting a Set of Straight Lines to a Digital BT Profile
ii. Author: W. E. Yergen
iii. TR-214

U.S. Naval Oceanographic Office
FITTING A SET OF STRAIGHT LINES TO A DIGITAL BT PROFILE, by W.E. Yergen, April 1969, 31 p. including figures and appendix. (TR-214)

A computer algorithm for fitting straight lines to subsets of points in a digital BT profile is described. A three level tolerance algorithm for selecting subsets of points that align is submitted, along with a method for determining an optimum thin line tolerance value. The algorithm may be applied to any kind of digital profile data set.

1. Algorithm, three level tolerance for fitting straight lines to point subsets in a digital profile.
2. Bathythermograph data reduction and representation.
3. Digital profile analysis algorithm.

i. title: Fitting a Set of Straight Lines to a Digital BT Profile
ii. Author: W. E. Yergen
iii. TR-214

U.S. Naval Oceanographic Office
FITTING A SET OF STRAIGHT LINES TO A DIGITAL BT PROFILE, by W.E. Yergen, April 1969, 31 p. including figures and appendix. (TR-214)

A computer algorithm for fitting straight lines to subsets of points in a digital BT profile is described. A three level tolerance algorithm for selecting subsets of points that align is submitted, along with a method for determining an optimum thin line tolerance value. The algorithm may be applied to any kind of digital profile data set.

1. Algorithm, three level tolerance for fitting straight lines to point subsets in a digital profile.
2. Bathythermograph data reduction and representation.
3. Digital profile analysis algorithm.

i. title: Fitting a Set of Straight Lines to a Digital BT Profile
ii. Author: W. E. Yergen
iii. TR-214

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

U. S. NAVAL OCEANOGRAPHIC OFFICE

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

FITTING A SET OF STRAIGHT LINES TO A DIGITAL BT PROFILE

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Walter E. Yergen

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1969	35	0

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. HF 05-243-301	TR-214
c. 707-FC-DAR	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

All distribution of this publication is controlled. All qualified DDC users shall submit requests for copies to the: Commander, U. S. Naval Oceanographic Office, Washington, D. C. 20390 Attn: Code 40

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	U. S. Naval Oceanographic Office

13. ABSTRACT

A "three level tolerance algorithm" for fitting sets of straight lines to points digitized at regular depth intervals from a BT profile is discussed. The application of this algorithm to the reduction of other profile data similarly digitized is discussed.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
FITTING A SET OF STRIGHT LINES TO A DIGITAL BT PROFILE						